Working Paper 3.3
Feb. 15, 2008

ICES WGMG REPORT 2007
ICES RESOURCE MANAGEMENT COMMITTEE
ICES CM 2007/RMC:04
Ref. ACFM

REPORT OF THE WORKING GROUP ON METHODS OF FISH STOCK ASSESSMENTS (WGMG)

(Excerpts from the above report are included here for consideration at the GARM-III meeting.)

13–22 MARCH 2007
WOODS HOLE, USA

ICES WGMG Report 2007 | 1 Executive summary

The purpose of the Working Group on Methods of Fish Stock Assessments (WGMG) is to develop and critically evaluate the models and software code used in assessments, forecasts and management simulations, and to suggest ways in which these might be improved. WGMG meets to address particular concerns raised by ACFM and the Resource Management Committee of ICES. The issues covered by each meeting are a function both of the Terms of Reference, and of the interests and expertise of the participants.

The 2007 meeting of WGMG was held at the Northeast Fisheries Science Center (NEFSC), NOAA, Woods Hole, USA. The principal reason for this was to draw on existing expertise at NEFSC on detecting and accounting for retrospective bias in fish stock assessments. The ToRs for the meeting were extremely wide, and covered many problems currently encountered in fisheries assessment and management science. With the time available WGMG could not address all the ToRs, so following an opening series of presentations of previous and current work, the group was divided into three subgroups to work on more focussed issues.

Subgroup A looked at methods for running management strategy evaluations (MSEs), and started designing simulations to assess how management advice might be affected by errors in assessments (in particular, retrospective bias). Subgroup B investigated ways in which the uncertainty in outputs from assessment models could be estimated. As a starting point, this was done by comparing Bayesian and bootstrap estimates of uncertainty arising from a comparatively simple surplus production model. Subgroup C looked further into the problem of retrospective assessment bias; that is, where each successive annual assessment substantially alters the perception of historical stock in a systematic way (either consistently increasing or decreasing it).

Subgroup A reached three main conclusions. Firstly, WGMG is not yet in a position to answer the questions of whether and how management should proceed in the presence of retrospective bias. The presence of such bias should lead to more cautious management, but how to implement this and how cautious such management should be is less clear. This is due principally to the complexity of programming management-strategy evaluations, but answers to these questions are certainly feasible using current approaches. Secondly, any management-strategy evaluation toolbox must allow for assessments to be run "live" as part of the evaluation loop. And thirdly, managers will get management decisions wrong if these are based on biased advice. This last point may seem obvious, but the analyses presented by Subgroup A highlights the issue with great clarity.

Assessments will always have problems of one sort or another, and it is important that MSEs are able to accommodate this fact. The function of WGMG in this regard is then to provide methods to do this. This endeavour therefore links the work of all three Subgroups.

Subgroup B provided important advances in the implementation of MCMC algorithms for model fitting, and went a considerable distance in generating comparisons of uncertainty estimates from bootstrap and Bayesian methods, with observation and/or process error, using a number of different datasets with different problems (one-way trips, under-reporting and changes in survey catchability). They were able to explore the varying reactions of models to these situations, but firm conclusions remained elusive due to considerable problems in software coding. The Section should be viewed as a strong advance in a work-in-progress. Nonetheless, it seems that not accounting for process errors can lead to a biased view of the true uncertainty in stock estimates based on approximate populations' models. Reliable methods that account for process and measurement errors simultaneously in stock assessment models are not yet available.

Subgroup C used four different techniques to try and detect model mis-specification in six simulated datasets. The techniques were:

Pre-screening of data inputs to assessment models.
Local influence diagnostics (LIDs).
The ADAPT approach with year effects in survey catchability (SPA YE).
The ADAPT approach with year effects in a catch multiplier (B-ADAPT)

In the case of LIDs, the method was used to try and ascertain the cause of retrospective bias *directly*; the use of the other methods was restricted to an evaluation of which model mis-specification had been applied (and when), without a concomitant analysis of the effect on retrospective bias (although this would be the next step). Pre-screening techniques can only be used to identify large changes in survey catchability. Similarly, the SPA YE model can only improve assessments when mis-specification of survey catchability is known to be the problem; and the B-ADAPT model performs best when errors in catch are the true source of mis-specification. LIDs suggested that survey catchability changes were responsible for retrospective bias in *all* simulations, even those in which the true cause was under-reporting and/or changes in natural mortality. In addition, correcting assessments using LIDs often removed retrospective bias but resulted in an incorrect assessment. These LIDs cannot therefore be considered reliable indicators for such problems, although they may still have utility when the VPA mis-specification is known to be small. However, a more positive result was that the diagnostics could more reliably detect the timing and direction of the problem when the source was known (e.g. M or survey catchability), especially in the more converged part of the VPA, Such models and diagnostics will perform best when used in combination with a) each other, and b) (more importantly) external information about the likely source of mis-specification.

Finally, analyses of different approaches to calculating a representative average *F* estimate for a given year were not able to determine any particular method that consistently performed well. Sensitivity of management advice to the method used needs to be evaluated on a case-by-case basis.

The main recommendations from the 2007 meeting of WGMG are summarised above. Of most direct relevance to this year's assessment Working Groups are the conclusions from Subgroup C, regarding testing for and correcting retrospective bias. The work of the other two Subgroups is at an earlier stage, but strong foundations for further work have been laid and plans are in train to continue collaborations. In addition, it was agreed that WGMG was an appropriate forum within which to carry forward certain aspects of size-based analyses; specifically, an exploration of the biases inherent in assuming size-based processes are age-based.

(pages 3-19 are deleted intentionally)

2.15 WP15

Chris Legault. Treatment of null values.

2.15.1 Abstract

Many stock assessment models use time series with assumed lognormal error that can have an observation of zero. For example, consider a research survey which does not encounter any age 1 individuals in 1995. This does not mean that there were no fish in that cohort, but rather that abundance, availability, and gear selectivity combined to produce an observation of zero. The standard procedure at the Northeast Fisheries Science Center (NEFSC) is to treat these null values as missing in stock assessments. Others advocate replacing the null with a small positive value *c*. Often $c = 1$, $c = 0.01$, or a rule is used which sets *c* equal to one sixth the smallest non-zero value in the series and adds this value to the entire time series. Treating these null values as missing avoids the issue of which value to use for *c* but could potentially bias the assessment by not providing information to the model. When there are multiple null values in a time series, replacing them all with a single value provides not just information about magnitude but also trend, which may or may not match the true trend in the population.

Simulations were conducted with survey values below an arbitrary level set to zero and these zeros then treated as missing, replaced with 0.01, or replaced using the one sixth rule. The $c = 0.01$ case performed poorly with highly biased estimates of *N* and *F*. Treating the zeros as missing performed best, with the one-sixth rule performing only slightly worse. A second set of simulations was conducted which added a second set of tuning indices which did not follow the true population trend under the assumption that this would be a harder test for the missing case. The results from this second set of simulations were hard to interpret. Treating the values as missing produced results closer to the case when all the values were used, but further from the underlying truth than the one-sixth case. The $c = 0.01$ case was quite highly biased for many more years than the missing case or one-sixth case. Given the inclusion of the biased indices, it is not clear whether the results using all the data or the underlying truth should be the basis for comparison.

The solution to the problem appears to be use the of a different error structure that allows for null values. Simulation testing will be required to demonstrate that such an alternative error structure is robust to outliers.

2.15.2 Summary of discussion

WGMG considered the use of replacing values with probabilities as part of the observation equation. Delta approaches were suggested as a possibility, whereby the probability of positive values is estimated separately from the magnitude of the positive values. Noel Cadigan mentioned that he encountered problems when trying this approach. Noel recalled the application of a quasi-likelihood function with a quadratic term as an approach to potentially reconcile this problem. This approach allows for null values and is close to the lognormal error distribution for positive observations. However, this approach requires the estimation of an additional parameter. The standard procedure in ICES is to treat null values in survey time series as missing. WGMG concluded that there is no simple solution to this problem, owing to the fact the log-normal distribution is an inappropriate error structure for these types of data series. WGMG stated that one should not change data to fit the model, but rather change the model to fit the data.

(remainder of report not included.)